A.1 Probability Review - Outline

Sums of Random Variables
- Independent Case
- Covariance
- Mixtures vs Sums
- Exercises

Sums of Independent Variables

For any random variables,

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$$
$$\mathrm{E}\left[\sum X_i\right] = \sum \mathrm{E}[X_i]$$

If in addition the variables are independent then

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$
$$\mathrm{Var}\left[\sum X_i\right] = \sum \mathrm{Var}[X_i]$$

What if X and Y are not independent?

Covariance

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \\
\mathrm{Cov}[X,Y] &= \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \mathrm{E}[XY] - \mathrm{E}[X]\cdot\mathrm{E}[Y] \\
\mathrm{Var}[X+Y] &= \mathrm{E}[(X+Y)^2] - (\mathrm{E}[X+Y])^2 \\
&= \mathrm{E}[X^2+2XY+Y^2] - \big((\mathrm{E}[X])^2 + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[Y])^2\big) \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 + 2(\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]) \\
&\quad + \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 \\
&= \mathrm{Var}[X] + 2\,\mathrm{Cov}[X,Y] + \mathrm{Var}[Y]
\end{aligned}$$

If X and Y are independent, then $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ and $\mathrm{Cov}[X,Y] = 0$. So if X and Y are independent

$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

Properties of Covariance

Covariance is a “bilinear form” meaning that it follows the usual distributive laws.

$$\begin{aligned}
(aX+bY)(cZ+dW) &= acXZ + adXW + bcYZ + bdYW \\
\mathrm{Cov}(aX+bY, cZ+dW) &= ac\,\mathrm{Cov}(X,Z) + ad\mathrm{Cov}(X,W) \\
&\quad + bc\,\mathrm{Cov}(Y,Z) + bd\mathrm{Cov}(Y,W) \\
\mathrm{Cov}(X,X) &= \mathrm{E}[X\cdot X] - \mathrm{E}[X]\cdot\mathrm{E}[X] = \mathrm{Var}[X] \\
(aX+bY)^2 &= a^2X^2 + 2abXY + b^2Y^2 \\
\mathrm{Var}[aX+bY] &= a^2\mathrm{Var}[X] + 2ab\,\mathrm{Cov}(X,Y) + b^2\mathrm{Var}[Y]
\end{aligned}$$

Recall the correlation of X and Y is

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{SD}[X]\mathrm{SD}[Y]}$$

$$-1 \le \mathrm{Corr}(X,Y) \le 1$$

Mixtures vs Sums

Example

A homeowners policy covers flood and theft losses. The amount of each flood loss is exponential with mean 10, the amount of each theft loss is exponential with mean 20, and a homeowner suffers one flood loss and one theft loss.
If the loss amounts are independent, and the insurance payment is 40% of the flood loss and 60% of the theft loss, what is the mean and variance of the payment?

Let $F \sim \exp(10) =$ flood losses, and $T \sim \exp(20) =$ theft losses.
The payment is $0.4F + 0.6T$, and

$$\mathrm{E}[0.4F + 0.6T] = 0.4 \cdot 10 + 0.6 \cdot 20 = 16$$

$$\mathrm{Var}[0.4F + 0.6T] = 0.4^2 \cdot 10^2 + 0.6^2 \cdot 20^2 = 160$$

This was a sum of two random variables, not a mixture.

Example

A homeowners policy covers flood and theft losses. The amount of each flood loss is exponential with mean 10 and the amount of each theft loss is exponential with mean 20. If 40% of losses are flood losses, and 60% are theft losses, what is the mean and variance of a randomly selected loss amount?

This is a mixture because a randomly selected loss is either a flood loss, or a theft loss, but not both at the same time.
Let X denote the loss amount.

Mixture example

With probability 0.4, $X \sim \exp(10)$,
and with probability 0.6, $X \sim \exp(20)$

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid \text{Case}]] \\
&= 0.4 \cdot 10 + 0.6 \cdot 20 = 16 \\
\mathrm{Var}[X] &= \mathrm{E}[\mathrm{Var}(X \mid \text{Case})] + \mathrm{Var}[\mathrm{E}(X \mid \text{Case})] \\
&= 0.4 \cdot 100 + 0.6 \cdot 400 + (20 - 10)^2 \cdot 0.4 \cdot 0.6 \\
&= 304 \\
\mathrm{P}[X \le x] &= 0.4\mathrm{P}[F \le x] + 0.6\mathrm{P}[T \le x] \\
&= 0.4\left(1 - e^{-x/10}\right) + 0.6\left(1 - e^{-x/20}\right) \\
&= 1 - 0.4e^{-x/10} - 0.6e^{-x/20}
\end{aligned}$$

Remark: Unlike sums, mixtures have nice densities and CDFs
But they have harder to find variances

Exercise 1

$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{Var}[X] = 1, \mathrm{Var}[Y] = 5, \mathrm{Var}[X + 2Y] = 13$. Find $\mathrm{E}[XY]$.

Exercise 1

$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{Var}[X] = 1, \mathrm{Var}[Y] = 5, \mathrm{Var}[X + 2Y] = 13$. Find $\mathrm{E}[XY]$.

$$
\begin{aligned}
\mathrm{Var}[X+2Y] &= \mathrm{Var}[X] + 2\mathrm{Cov}(X, 2Y) + \mathrm{Var}[2Y] \\
&= \mathrm{Var}[X] + 4\mathrm{Cov}(X, Y) + 4\mathrm{Var}[Y] \\
13 &= 1 + 4 \cdot \mathrm{Cov}(X, Y) + 4 \cdot 5 \\
\mathrm{Cov}(X, Y) &= -2 \\
\mathrm{Cov}(X, Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \\
-2 &= \mathrm{E}[XY] - 2 \cdot 3 \\
\boxed{4} &= \mathrm{E}[XY]
\end{aligned}
$$

Exercise 2

Let X and Y be independent variables with $\mathrm{E}[X] = 3$, $\mathrm{E}[Y] = 5$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = 4$. Suppose that $W = 0.4X + 0.6Y$, and let Z be a variable that equals X 40% of the time and Y 60% of the time. Find $\mathrm{E}[W], \mathrm{E}[Z], \mathrm{Var}[W]$ and $\mathrm{Var}[Z]$.

Exercise 2

Let X and Y be independent variables with $\mathrm{E}[X] = 3$, $\mathrm{E}[Y] = 5$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = 4$. Suppose that $W = 0.4X + 0.6Y$, and let Z be a variable that equals X 40% of the time and Y 60% of the time. Find $\mathrm{E}[W], \mathrm{E}[Z], \mathrm{Var}[W]$ and $\mathrm{Var}[Z]$.

Remark: W is a sum, Z is a mixture.

$$\begin{aligned}
\mathrm{E}[W] &= \mathrm{E}[0.4X + 0.6Y] = 0.4\mathrm{E}[X] + 0.6\mathrm{E}[Y] = 4.2 \\
\mathrm{E}[Z] &= 0.4\mathrm{E}[X] + 0.6\mathrm{E}[Y] = 4.2 \\
\mathrm{Var}[W] &= 0.4^2\mathrm{Var}[X] + 2 \cdot 0.4 \cdot 0.6\mathrm{Cov}(X, Y) + 0.6^2\mathrm{Var}[Y] \\
&= 0.16 \cdot 4 + 0 + 0.36 \cdot 4 = 2.08 \\
\mathrm{Var}[Z] &= \mathrm{E}[\mathrm{Var}[Z \mid \mathrm{Case}]] + \mathrm{Var}[\mathrm{E}[Z \mid \mathrm{Case}]] \\
&= 0.4 \cdot 4 + 0.6 \cdot 4 + (5-3)^2 \cdot 0.4 \cdot 0.6 \\
&= 4.96
\end{aligned}$$